DOCUMENT RESUME

ED 222 566 TM 820 724

AUTHOR Roth, Rodney
TITLE Validation of the NTE: Arkansas Style.
PUB DATE Jun 82
NOTE 16p.; Paper presented at the Education Commission of the States Annual Conference on Large Scale Assessment (Boulder, CO, June 10, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cutting Scores; Elementary Secondary Education; Item Analysis; Public Schools; *State Programs; State Standards; *Teacher Certification; Teacher Evaluation; Teaching Skills; *Testing Programs; *Test Selection; Test Use; Test Validity
IDENTIFIERS *Arkansas; Arkansas State Department of Education; *National Teacher Examinations

ABSTRACT

The National Teacher Examination (NTE) validation and cut score determination procedures for use in teacher certification in Arkansas are discussed. The NTE area examinations to be validated were selected by the Arkansas Department of Education, which compared the state's certification requirements with the NTE's rationale, content, and scope descriptions. The 24 examinations selected were in such areas as art education, business education, the sciences, several languages, early childhood and elementary education, guidance and counseling, and other subject specializations. The Educational Administration and Supervision Examination was also selected. Judges selected for the NTE Review Study Group included certified practitioners from public schools, colleges, and universities in the state. Unique procedures used in the Arkansas study included assessing examination item validity by item relevance and providing national p-values to assist judges in finding item probabilities for determining cut-scores. The legal rationale for using the procedures is discussed. (Author/CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED222566

VALIDATION OF THE NTE: ARKANSAS STYLE

ABSTRACT

This paper presents the NTE validation and cut-score determination procedures used by the author for the State of Arkansas in order for Arkansas to use the NTE for teacher certification.

Several unique procedures were used in the Arkansas study. Some of these were to have public school practitioners serve as judges, item validity was assessed by item relevance, and national p-values were provided to assist judges in judging item probabilities for cut-score determination.

The paper concludes with a legal rationale for using the procedures in the Arkansas study.

TM 820 224

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Roth

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

VALIDATION OF THE NTE: ARKANSAS STYLE*

Rodney Roth, Director
Bureau of Field Studies and School Services
University of Arkansas

Introduction

The Arkansas Legislature passed Act 162, in March, 1979. The Act stated that after January 1, 1980, any person applying for initial certification as a teacher in the public schools or a certified teacher applying for certification in an additional area shall, in addition to other requirements, take and complete the National Teacher's Examination or a similar examination designated by the State Board of Education and submit the scores to the Department of Education.

A positive aspect of Act 162 was that a cut-score did not have to be determined until January, 1982. A negative aspect, but in retrospect it was positive, was that the legislature did not appropriate to the State Department any money to determine a cut score and/or develop a test.

Time and money problems prompted the State Board of Education to require the NTE for certification purposes after January 1, 1980.

The State Department of Education appointed a NTE Review Study Group to establish a process of validation and cut-score determination during the Fall, 1980. This group was informed by ETS that the NTE Commons was being revised and would be available in November, 1982. This group recommended that Act 162 be changed during the 1981 legislative session because it would be a waste of money to validate an "old" test.

---

* Paper presented at the Education Commission of the States Annual Conference on Large-Scale Assessment, Boulder, June 10, 1982.

The Arkansas Legislature did change the law (They meet once every two years.) in March, 1981. The change moved the cut-score date to February 1, 1983 and allocated some money to the Department of Education for validation.

The NTE Review Study Group, in the Spring, 1981, recommended that the area exams be validated in Arkansas. They also recommended that the Department of Education distribute a RFP for the NTE validation and cut-score study. During the Fall, 1981, the Bureau of Field Studies and School Services, University of Arkansas was awarded the NTE study contract through the bid process.

NTE Study Design

This section of the paper presents the NTE area exams selected for the study, judge selection procedures and the data collection procedures used for the study.

NTE Areas Studied

The Arkansas Department of Education selected the NTE area exams to be validated. The decision was based on comparing the Arkansas certification requirements and the NTE rationale, content and scope descriptions published by ETS for each area exam. The area exams selected were: Art Education; Business Education; Biology and General Science; Chemistry, Physics, and General Science; Early Childhood Education; Education in the Elementary School; English; French; German; Guidance and Counseling; Home Economics Education; Industrial Education; Introduction to the Teaching of Reading; Mathematics Education; Media Specialist; Music Education; Physical Education; Reading Specialist; Social Studies; Spanish; Speech Communications; and Speech Pathology.

The German exam was not validated due to a lack of judges in Arkansas. The Speech Communications exam was not validated since ETS is introducing a new test for this area in February, 1983. The Mental Retardation test was not selected for validation since Mental Retardation will not be a certifiable area in Arkansas after September, 1984.

The area exam in Educational Administration and Supervision was also selected for validation. In addition, three different cut-scores for certification will be derived from this one test. These will be for the Elementary School Principal's Certificate, Secondary School Principal's Certificate and the Administrator's Certification. The rationale for this decision was that the training programs are 30 graduate hours, 45 graduate hours and 60 graduate hours, respectively.

Selection of Judges

The Arkansas Department of Education decided that the judges for the NTE study should include certified practitioners from Arkansas public schools and faculty members from colleges and universities in Arkansas.

Each College of Education dean in the sixteen teacher training institutions was asked to nominate judges from his/her institution for the NTE study. Each dean was asked to nominate judges only in the NTE areas in which the college had approved certification programs. The nominated judges filled out a nomination form which included information about race, sex, years of teaching experience and courses taught. These forms were sent to the project director at the University of Arkansas.

The nominations for NTE judges from practitioners were made by fourteen superintendents from throughout Arkansas. Each superintendent was asked to nominate a certified person for each NTE area to be validated. In addition, they recommended elementary and secondary principals. The

nominated practitioners also sent nomination forms to the project director. These forms were similar to the college faculty forms.

The superintendents selected for the NTE study were selected from nominations supplied by the Executive Director of the Arkansas Association of Educational Administrators.

The actual selection of the final set of judges was made by the project director from the pool of nominations made by deans and superintendents. A panel of judges was selected for each NTE area exam. Each panel consisted of 14-16 persons with approximately 7 or 8 public school practitioners and 7 or 8 faculty members from colleges and universities..

Sone of the criteria used to select the judges were race, sex, years of teaching/administrative experience, teaching assignment, and for college faculty, the number of graduates produced by their institution.

A total of 333 judges participated in the NTE study. One hundred sixty one were associated with public schools in Arkansas and 172 were from colleges and universities.

Data Collection

The data for the study were collected between April 20, 1982 - April 28, 1982, at five different locations in Arkansas. At each location, the judges from colleges and universities met from 12:30 - 3:30 p.m. and the public school judges met from 6:30 - 9:30 p.m.

Each data collection session began with a training session. It included a legal history of the NTE in Arkansas, purpose of the NTE area exams, the need for state validation, and the NTE study design including how the judges were selected. The training session also included very specific directions for the validity and cut score judgments. These directions were:

The first rating you will make concerns item Relevance. This will be used for test validation. In order to make this judgment, you should read the item, the "correct" answer, and the distractors. (The correct answer is underlined in the test booklet.) You should then judge the relevance of the content measured by the question with respect to the domain of knowledge you believe a minimally qualified entry-level person in the certification area should possess.

If you believe the content of the question is irrelevant to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 1 on your answer sheet in the Relevance column to signify "Not Relevant."

If you believe the content of the question is of doubtful or questionable relevance to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 2 on your answer sheet in the Relevance column to signify "Questionable."

If you believe the content of the question is important, but not quite crucial, to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 3 on your answer sheet in the Relevance column to signify "Important."

If you believe the content of the question is of crucial importance to the domain of knowledge a minimally qualified entry-level person in this field should possess, then you should fill-in circle 4 on your answer sheet in the Relevance column to signify "Crucial."

Points to consider for the relevance ratings:

A. A minimally qualified entry-level person will have completed the training requirements as specified by Arkansas Certification Requirements. The requirements were sent to you for your NTE area.

B. A minimally qualified entry-level person has the minimal amount of knowledge to initially perform in the certification area.

C. In making your relevance judgment, be sure to think in terms of the full range of grade levels and/or subject areas and/or types of schools for your certification. Examples:

   a. If you are judging the NTE Elementary Education for Elementary (1-6) Certification and you find an

item "Not Relevant" for early elementary grade teachers, but "Crucial" for later elementary grade teachers, then you would mark the item "Crucial."

b. If you are judging the NTE Math exam for Math Certification and you find an item "Questionable" for teaching 9th or 10th grade math, but "Important" for teaching 11th or 12th grade math, then you would mark the item "Important."

c. If you are judging for the Elementary Principal Certification and find an item "Crucial" for a large urban elementary school, but "Questionable" for a small town elementary school, then you should mark the item "Crucial."

d. If you are judging the NTE in Biology and General Science, then you should respond for Biology or General Science Certification. If you are judging the NTE in Chemistry and Physics, then you should respond for Chemistry or Physics or Physical Science Certification.

The second judgment you will make about each item will help determine the cut-score. You should imagine a hypothetical person, who in your judgment, has the minimum amount of academic knowledge to complete the preparation program required for certification in Arkansas and has the minimum amount of knowledge to perform in the field designated by the NTE area test. With this hypothetical person in mind, you are to estimate the probability that this minimally competent person would know the answer to the NTE item without guessing. Another way of thinking about this estimation process is to think of a group of minimally competent persons and then estimate the percent of minimally competent persons who would answer the NTE item correctly without guessing.

Before you make your estimate about the item, you should also realize the item difficulty based on the NTE norm group for the item. The item difficulty or the percent who have passed the item is written beside the item in the booklet.

You should mark your estimate for each item on the response sheet under the Probability column. You should use the following scale for these estimates:

Fill in circle 1, if your estimate is between .00 - .10
Fill in circle 2, if your estimate is between .11 - .20
Fill in circle 3, if your estimate is between .21 - .30
Fill in circle 4, if your estimate is between .31 - .40
Fill in circle 5, if your estimate is between .41 - .50

Fill in circle 6, if your estimate is between .51 - .60
Fill in circle 7, if your estimate is between .61 - .70
Fill in circle 8, if your estimate is between .71 - .80
Fill in circle 9, if your estimate is between .81 - .90
Fill in circle 10, if your estimate is between .91 - 1.0

The judges were able to make the two judgments per item in less than three hours. Most NTE exams have 150 items. One, however, does have 180 items. The actual responses by the judges were made on a specifically designed NCS answer sheet which was machine scanned.

Data Analysis

This section of the paper presents the methods used to determine item validity and cut scores.

Validity

The validity of each item was determined by computing an item mean for each item on the relevance scale. This scale had a range from one (Not Relevant) to four (Crucial). In order for an item to be considered valid, the mean score on the relevance scale had to be greater than 2.5. In other words, the item had to be rated by the judges as closer to the important category than to the questionable category. If half of the judges had rated the item questionable and the other half had rated the item important, then the item would not have met the validity criterion since the mean rating would have been 2.50.

Cut-Scores

The cut-score for each area exam was determined by a slight modification of a procedure known as the Angoff method. The first step for determining the cut-score was to determine an item mean on the probability scale. Since the judges had responded to a probability range for each item, the mid-point of the range was used to compute the item mean. For

example, a one on the probability scale was converted to .05 since one represented the probability between .0 thru .1.

The raw score cut-score for each area exam was computed by summing the mean probabilities for only the items that had met the validity criterion. A conversion foumula can then be used to convert the raw scores to NTE standard scores.

Legal Rationale*

This section of the paper presents the legal rationale for using the validation procedure that was used in Arkansas. In order to build the rationale, the NTE decision in South Carolina, ETS validation procedures and the Uniform Guidelines will be briefly reviewed.

South Carolina Decision

On January 16, 1978, the United States Supreme Court announced it had summarily affirmed the April, 1977 decision of the Federal District Court upholding South Carolina's use of the NTE in the State's certification system.

Five justices affirmed the decision of the Federal District Court of South Carolina concerning use of the NTE for certification. They did not, however, issue an opinion for their reasoning. Justices White and Brennan did, however, record a strong dissent. Furthermore, the case has not been cited as authority in any other Federal or State court opinion with regards to the use of NTE for certification purposes.

---

* The legal rationale is based on a monograph published by the Bureau of Field Studies and School Services, University of Arkansas. It is entitled Testing for Teacher Certification - Legal and Technical Considerations. The authors are Ms. Sharon Streett, Attorney, Fayetteville, Arkansas and Dr. Rod Roth, Director, Bureau of Field Studies and School Services.

Some of the reasons the South Carolina Court used to enable the State to use the NTE were:

1. There was no intent to discriminate against blacks in developing the testing system and that rather than a racial classification, the system created a permissable classification based on knowledge or the lack of knowledge;

2. The State's interest in hiring at least minimally competent teachers was related in a rational manner to its use of the NTE scores.

3. The use of NTE scores was a valid "business necessity." This means that in some instances a discriminatory practice might be tolerated if it were essential to the operation of the business. The "business necessity" rationale used by the South Carolina Court was based on a Supreme Court decision which said:

> The Supreme Court intended on examination of the alternatives available with respect to the legitimate employment objective identified by the employer to determine whether there is available to the employer an alternative practice that would achieve his business purposes equally well but with a lesser disparate impact by race. In examining alternatives, the risk and cost to the employer are relevant. (Griggs v. Duke Power Co.)

In the United States v. South Carolina case, the Plaintiffs presented only one alternative, graduation from an approved teacher training program, to the use of the NTE for certification purposes. The trial Court did not feel that the alternative would achieve the State's purpose in certifying minimally competent teachers as well as the use of the NTE. The Court in support of this finding made two points. One, evidence demonstrated that the teacher training programs varied in admission requirements, academic standards, and grading practices. Two,

evidence demonstrated that the State approves only general subject matter areas covered by the programs, not the actual course content of the programs. Both of these points would seem to weigh negatively on the Court's position that validation against the teacher training programs was sufficiently reflective of actual knowledge needed for the teaching positions. Here the Court would seem to be admitting that the twenty-five teacher training programs were in fact different and therefore not all would be to the same degree reflective of knowledge needed to competently perform the job. The Court, however, while finding the teaching programs themselves an inadequate measure of teacher competency saw no inconsistency in finding test validation against those same teacher programs acceptable.

ETS Validation Procedures

ETS (1976) in their Brief Amicus Curiae filed in the District Court of South Carolina stated:

> Any general challenge by the plaintiffs to the use of the NTE by South Carolina for measuring the "academic preparation" of prospective teachers would be foreclosed by the North Carolina decision. There the three-judge court held that:
>
> > beyond argument . . . the State of North Carolina has the right to adopt academic requirements and written achievement tests designed and validated to disclose the minimum amount of knowledge necessary to effective teaching. 400 F. Supp. at 348. (p. 4)

All of the states that have a cut-score on the NTE for certification have hired ETS to do their validity studies. The validity studies in these states followed very similar procedures with review panels selected from teacher training institutions in the states. ETS (1976) describes the process as follows:

> The Content Review Panels were asked to make four independent types of judgments about the congruence of the NTE tests they were reviewing and the curriculum in teacher training programs in South Carolina. Judges were asked to indicate,

on the basis of a review of the major content topics of the test, whether the emphasis in the curriculum on each content topic was about the same as, more than, or less than the percentage of the test devoted to the content topic. As an additional means of assessing the overall similarity between tests and curricula, judges were also asked to indicate one of four characterizations of greater or lesser similarity between the test and the curriculum of their institution. As a measure of the comprehensiveness of the tests, judges were asked to indicate which, if any, major content topics in the teacher education sequence were not included among the test content topics. Finally, the judges were asked to review each question on the test to determine whether the individual question was appropriate for use in South Carolina. The questions on the NTE were regarded as "content appropriate" if at least 51% of the judges indicated that at least 90% of the students taking the test would have had an opportunity to learn the answer as part of their teacher education program or in courses prerequisite to entering it. (pp. 17-18)

This content validation process was upheld by courts in North Carolina and South Carolina when they ruled in favor of using the NTE for certification in these States.

Uniform Guidelines

An important point concerning the South Carolina decision is that the decision was made before the Uniform Guidelines on Employment Selection Procedures were adopted in 1978 by the Equal Employment Opportunity Commission, the Civil Service Commission, the Department of Labor, the Department of Justice and Treasury. These Guidelines are intended to assist private employees, government contractors, employment agencies, labor organizations, state and local governments and the Federal Government in complying with Federal law prohibiting discrimination by race, color, religion, sex and national origin. Employment decisions such as hiring, promotion, demotion, membership, referral, retention, licensing and certification are encompassed within the scope of the Guidelines.

The Uniform Guidelines do not require validation studies where a selection method does not adversely affect (An adverse effect is a

selection rate for any racial, ethnic or sex group which is less than 80% of the rate for the highest scoring group.) a protected group. Where validation is required, however, content, construct and criterion-related validity studies are equally acceptable.

Section 5, Uniform Guidelines on Employee Selection Procedures, General Standards for Validity Studies states:

> Evidence of the validity of a test or other selection procedure by a criterion related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the selection procedure is a representative sample of important work behavior to be performed on the job for which the candidates are to be evaluated. Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important is successful performance in the job for which the candidates are to be evaluated.

The underlining was done by this writer to indicate that regardless of the type of validity study one uses, the test one uses for certification has to be directly job related. Furthermore, Section 14, Uniform Guidelines on Employee Selection Procedures, Technical Standards for Content Validity states:

> Where a training program is used as a selection procedure and the content of a training program is justified on the basis of content validity, it should be justified on the relationship between the content of the training program and the content of the job.

In reference to the above guideline section, Paul Tractenberg (1980), Professor of Law, Rutgers University states:

> The South Carolina NTE case does seem to suggest that content validity may be sufficient and does seem to suggest even that content validity may be sufficient if it is related

to training programs; that is, the test items connect up to the training programs rather than to the job on the other end. That may be the only court decision that says really quite that, and as such, it is both a reasurrance and a matter of some concern that it will become the uniform rule of the courts, especially in light of the new Uniform Guidelines on Employee Selection Procedures which have a provision that seems to link content validity to what happens on the job, not simply what happened in the training program.

Legal Rationale for Arkansas Procedures

A paramount issue, in addition to measurement concerns, in the Arkansas NTE study has been to make the procedures and the final results as legally defensible as possible.

The project director, after consulting with several attorneys, believes that the ETS validation procedures used in the other "NTE States" would not be upheld in a current court case. The primary reasons for this belief are the Uniform Guidelines on Employee Selection Procedures and court decisions made after the South Carolina NTE case.

In Ensley v. Branch of N.A.A.C.P. v. Seibels, 616 F.2d 812 (5th. Cir. 1980) there is a broad language indicating that validation to a training program is insufficient. The court says:

> "We do not believe the Davis rationale can be extended... to the general proposition that any test can be validated by showing a relationship to training."
>
> "We decline to extend the Davis rationale by holding that any test can be validated against training, without respect to the test's ability to predict job performance. Such an extension would violate the requirement of job performance validation enunciation in Griggs and Abermarle as well as agency guidelines (Uniform Guidelines) elaborating upon that requirement.

A case which explicitly says that the EEOC guidelines should apply is United States v. Virginia, 620 F.2d 1018 (4th Cir. 1980). In Virginia the state was giving a test for employment of state police

officers. The state had validated the test according to the content of the training programs for the job. The court said:

> "In Davis the court held that a positive relationship between the test and training-course performance was a sufficient validation under standards similar to those of Title VII. While the evidence presented to the district court was sufficient to support its conclusion under Davis, upon remand the court should reconsider this issue in the light of the applicable EEOC guidelines to determine whether a contrary conclusion is required."

The reader should realize that the Davis case was used by the South Carolina Court in their NTE decision. Furthermore, the Davis case was also rendered before the Uniform Guidelines were published.

The Arkansas NTE validation procedure was designed to meet the intent of the general validity standards of the Uniform Guidelines. In other words, the validity question was designed to be job related, instead of training program related.

Ary state using the NTE may, however have legal problems, regardless of the validation procedures. A recent test validation case is Guardians Association of the New York City Police Department, Inc. v. Civil Service Commission, 630 F. 2d 79 (1980). The court in this case stated five steps for test validation. They were:

1. the test-makers must have conducted a suitable job analysis; and
2. they must have used reasonable competence in constructing the test itself;
3. the content of the test must be related to the content of the job;
4. the content of the test must be representative of the content of the job; and
5. a scoring system must be used which usefully selects from among the applicants those who can better perform the job.